SYSTEM AND ARCHITECTURE FOR PRIVACY-PRESERVING DATA MINING

ABSTRACT OF THE DISCLOSURE

A system and method for mining data while preserving a user's privacy includes perturbing user-related information at the user's computer and sending the perturbed data to a Web site. At the Web site, perturbed data from many users is aggregated, and from the distribution of the perturbed data, the distribution of the original data is reconstructed, although individual records cannot be reconstructed. Based on the reconstructed distribution, a decision tree classification model or a Naive Bayes classification model is developed, with the model then being provided back to the users, who can use the model on their individual data to generate classifications that are then sent back to the Web site such that the Web site can display a page appropriately configured for the user's classification. Or, the classification model need not be provided to users, but the Web site can use the model to, e.g., send search results and a ranking model to a user, with the ranking model being used at the user computer to rank the search results based on the user's individual classification data.